**Chung et al., 2014:**

@article{chung2014empirical,

title={Empirical evaluation of gated recurrent neural networks on sequence modeling},

author={Chung, Junyoung and Gulcehre, Caglar and Cho, KyungHyun and Bengio, Yoshua},

journal={arXiv preprint arXiv:1412.3555},

year={2014}

}

**Devlin et al., 2019:**

@inproceedings{devlin2019bert,

title={Bert: Pre-training of deep bidirectional transformers for language understanding},

author={Devlin, Jacob and Chang, Ming-Wei and Lee, Kenton and Toutanova, Kristina},

booktitle={Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)},

pages={4171--4186},

year={2019}

}

**Reimers & Gurevych, 2019**

@article{reimers2019sentence,

title={Sentence-bert: Sentence embeddings using siamese bert-networks},

author={Reimers, Nils and Gurevych, Iryna},

journal={arXiv preprint arXiv:1908.10084},

year={2019}

}

**Khandelwal et al. (2020)**

@article{khandelwal2020nearest,

title={Nearest neighbor machine translation},

author={Khandelwal, Urvashi and Fan, Angela and Jurafsky, Dan and Zettlemoyer, Luke and Lewis, Mike},

journal={arXiv preprint arXiv:2010.00710},

year={2020}

}